



Actividad 2

Alumno: José Osvaldo Farfán de León

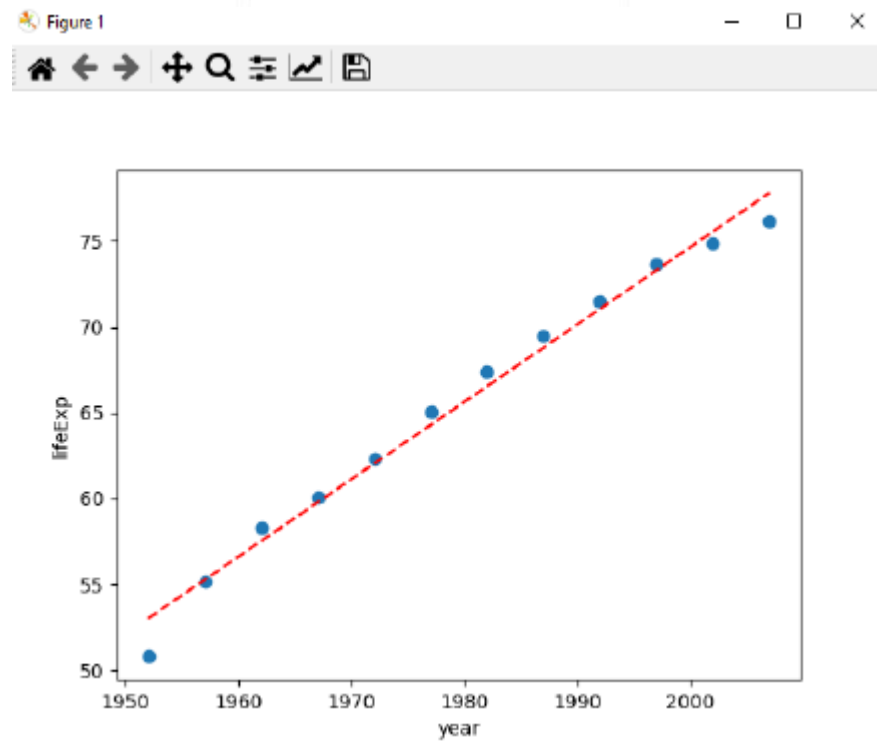
Materia: Seminario de IA II

Profesor: Carlos Alberto Villaseñor Padilla

Sección: "D04"

Crea una regresión lineal simple usando Scikit-learn usando el dataset "countries.csv" y realiza lo siguiente:

Crea un gráfico de dispersión del año contra la esperanza de vida en México.



Responde: La variable año y la variable esperanza de vida están correlacionadas?

No, ya que el año no nos proporciona información sobre el comportamiento de la esperanza de vida.

Haz regresión lineal con esos datos y has las siguientes predicciones:
cual es la esperanza de vida en México en:

2005

```
In [4]: model.predict([[2005]])
Out[4]: array([[76.91022494]])
```

La predicción para el 2005, nos dice que la esperanza de vida alcanzaba los 76 años

2019

```
In [5]: model.predict([[2019]])
Out[5]: array([[83.22471445]])
```

La prediccion para el 2019, nos dice que la esperanza de vida alcanzaba los 83 años

3019

```
In [6]: model.predict([[3019]])  
Out[6]: array([[534.25967949]])
```

La prediccion para el 3019, nos dice que la esperanza de vida alcanzaba los 534 años

-542

```
In [7]: model.predict([[ -541]])  
Out[7]: array([[ -1071.42479604]])
```

La prediccion para el -541, nos dice que la esperanza de vida alcanzaba los -1071 años

Explica el resultado de tus predicciones.

Al realizar una regresión lineal a 2 variables que no están correlacionadas, obtendremos resultados que pierden lógica, ya que entre mayor sea el año, mayor será la esperanza de vida y viceversa, por lo que podemos notar que en años como el ingresado -541 obtenemos un valor que es imposible de representar en nuestra realidad. De la misma manera que obtenemos un resultado ilógico al utilizar números negativos, si realizamos nuestra predicción con años aun mas grandes, obtendremos una esperanza de vida muy alta. Al no existir una correlación y ser un modelo de regresión lineal, entre mayor sea el año, mayor será la esperanza de vida sin justificar el porque el aumento de uno equivale al aumento del otro.

Explica la diferencia entre correspondencia y causalidad?

La correspondencia indica que ambos sucesos pueden crecer o decrecer de juntos existiendo o no una relacion entre ambos sucesos.

Mientras que la causalidad indica que existe una relacion entre los sucesos, donde un suceso influye al cambio en el cambio del otro suceso.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model, metrics

#LEEMOS EL ARCHIVO#
df = pd.read_csv('countries.csv')
#GUARDAMOS LOS REGISTROS DE MEXICO#
df_mex = df[df.country == 'Mexico']
#CONSTRUIMOS LA GRAFICA CON AÑO Y ESPERANZA DE VIDA#
df_mex.plot.scatter(x='year',y='lifeExp')

# MODELO REGRESION LINEAL #

# VARIABLES DE ENTRADA #
x = np.asanyarray(df_mex[['year']])

# VARIABLES DE SALIDA #
y = np.asanyarray(df_mex[['lifeExp']])

# INSTANCIAR EL MODELO #
model = linear_model.LinearRegression()
model.fit(x, y)

# PREDICCIONES #
y_pred = model.predict(x)

# GRAFICAR #
plt.scatter(x,y)
plt.plot(x,y_pred,'--r')
plt.show()

print("MAE: ",metrics.mean_absolute_error(y, y_pred))
print("MSE: ",metrics.mean_squared_error(y, y_pred))
print("MEDIANA DEL ERROR ABSOLUTO: ",metrics.median_absolute_error(y, y_pred))
print("R2: ",metrics.r2_score(y, y_pred))
print("EVS: ",metrics.explained_variance_score(y, y_pred))
```